

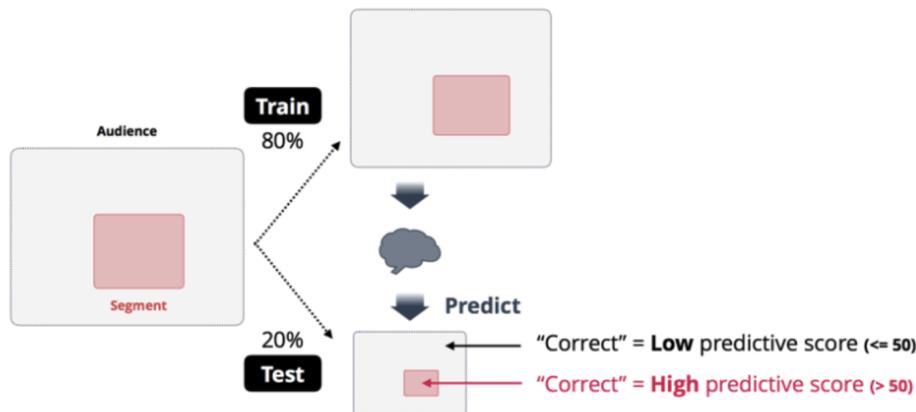
# About Training and Predictive Scoring Determination

 This article supports Audience Studio - Legacy.

When you train Predictive Scoring, an evaluation runs in parallel to estimate the accuracy of prediction. The evaluation completes the following steps:

1. Splits profiles in population into 80% training and 20% testing samples.
2. Builds a predictive model by using only the 80% train set.
3. Computes a predictive score for profiles in the 20% test set.
4. Puts an estimated "converted or not" label to each of the 20% profiles based on the predictive score. For example, if the predictive score is greater than 50, a profile is identified as "convert in near future" (becomes a positive sample).

Each test profile has a *predictive score* and *estimated label*. And, because positive samples is known, the *truth label* for the test profile is obvious. We can compute the accuracy of prediction by comparing the *truth label* to the *predictive score/estimated label*:



The metrics of Accuracy and AUC ([Area Under the ROC Curve](#)) are derived from *estimated label* and *predictive score* for the 20% test sample.

Accuracy is the percentage of the "correct" estimated label computed over pairs of truth label and estimated label:

truth label (in positive samples)	estimated label (predictive score > 50)	
1	0	Incorrect
0	0	Correct
1	1	Correct

Because a rounded or truncated estimated label is less informative than a raw predictive score, Accuracy is not reliable as a metric. Therefore, we recommend that you consider the AUC (shorthand for ) which is a metric computed from raw *predictive score*.

The metric returns a float value (larger is better) in [0.0, 1.0] range, and higher (lower) predictive score for *truth label* = 1 (0) customer increases AUC value, and vice versa.

Data sometimes generates poor prediction results. Exporting data for use by another system based on poor predictive results does not contribute to your business success. Treasure Data provides auxiliary information about the results of predictive scoring to help you refine your results.